BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e39229
doi: 10.3897/biss.3.39229

OPEN ACCESS

Conference Abstract

Machine Learning Model for Identifying Dutch/ Belgian Biodiversity

Laurens Hogeweg[‡,§], Maarten Schermer[‡], Sander Pieterse[‡], Timo Roeke[|], Wilfred Gerritsen[‡]

‡ Naturalis Biodiversity Center, Leiden, Netherlands
§ COSMONiO Imaging BV, Groningen, Netherlands
| Observation.org, Groningen, Netherlands

Corresponding author: Laurens Hogeweg (laurens.hogeweg@naturalis.nl)

Received: 19 Aug 2019 | Published: 20 Aug 2019

Citation: Hogeweg L, Schermer M, Pieterse S, Roeke T, Gerritsen W (2019) Machine Learning Model for Identifying Dutch/Belgian Biodiversity. Biodiversity Information Science and Standards 3: e39229.
https://doi.org/10.3897/biss.3.39229

Abstract

The potential of citizen scientists to contribute to information about occurrences of species and other biodiversity questions is large because of the ubiquitous presence of organisms and friendly nature of the subject. Online platforms that collect observations of species from the public have existed for several years now. They have seen a rapid growth recently, partly due to the widespread availability of mobile phones. These online platforms, and many scientific studies as well, suffer from a *taxonomic bias*: the effect that certain species groups are overrepresented in the data (Troudet et al. 2017). One of the reasons for this bias is that the accurate identification of species, by non-experts and experts, has been limited by the large number of species that exist. Even in the geographically limited area of the Netherlands and Belgium, the number of species that are regularly observed are in the thousands. This makes the ability to identify all those species difficult or impossible for an individual.

Recent advances in species identification powered by deep learning, based on images (Norouzzadeh et al. 2018), suggest a large potential for a new set of digital tools that can help the public (and experts) to identify species automatically. The online observation platform Observation.org has collected over 93 million occurrences in the Netherlands and Belgium in the last 15 years. About 20% of these occurrences are supported by photographs, giving a rich database of 17 million photographs covering all major species

© Hogeweg L et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

groups (e.g., birds, mammals, plants, insects, fungi). Most of the observations with photos were validated by human experts at Observation.org, creating a unique database suitable for machine learning. We have developed a deep learning-based species identification model using this database containing 13,767 species, 1,530 species-groups, 734 subspecies and 117 hybrids. The model is made available to the public through a web service (https://identify.biodiversityanalysis.nl) and through a set of mobile apps (ObsIdentify).

In this talk we will discuss our technical approach for dealing with the large number of species in a deep learning model. We will evaluate the results in terms of performance for different species groups and what this could mean to address part of the taxonomic bias. We will also consider limitations of (image-based) automated species identification and determine venues to further improve identification. We will illustrate how the web service and mobile apps are applied to support citizen scientists and the observation validation workflows at Observation.org. Finally, we will examine the potential of these methods to provide large scale automated analysis of biodiversity data.

Keywords

machine learning, deep learning, image recognition

Presenting author

Laurens Hogeweg

Presented at

Biodiversity_Next 2019

References

- Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer M, Packer C, Clune J (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences 115 (25): E5716-E5725. https://doi.org/10.1073/pnas.1719367115
- Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F (2017) Taxonomic bias in biodiversity data and societal preferences. Scientific Reports 7: 9132. https://doi.org/10.1038/s41598-017-09084-6